Performance Evaluation of CT Systems

AAPM Task Group 233
Foundations, Methods, and Prospects

Ehsan Samei
Duke University Medical Center

Disclosures

- Research grant: NIH R01 EB001838
- Research grant: General Electric
- Research grant: Siemens Medical
- Advisory board: Imologix

TG 233 membership

Donovan Bakalyar
Kirsten L Boedeker
Samuel Brad
Jiahua Fan
Shuai Leng
Michael McNitt-Gray
Kyle J. Myers
Lucretiu Popescu
Juan Carlos Ramirez Giraldo
Frank Ranallo
Ehsan Samei
Justin Solomon
Jay Vaishnav
Jia Wang

Duke Imaging Physics

Carl E Ravin Advanced Imaging Laboratories
Clinical Imaging Physics Group

Outline

- **Foundation**
  - **What has led to TG233**
- **Methods**
  - **Select procedures**
- **Prospects**
  - **Possibilities**
  - **Future extensions**

Outline

- **Foundation**
  - **What has led to TG233**
- **Methods**
  - **Select procedures**
- **Prospects**
  - **Possibilities**
  - **Future extensions**

Foundation

- New physics perspective
- New CT features

New physics perspective

- Medical physics metrology is most relevant to the extend it relates to the clinical need

1. Conformance-based testing
   - Validating physical specifications: pass/fail
   - Relevance of non-compliance?
2. Performance-based testing
   - Metrics relevant to clinical accuracy
   - Application to optimization of use: precision medicine and patient-centric care

Medical Physics 3.0

Medical physics extending from

specifications	to	performance
equipment	to	operation
quality check	to	process consistency
Presumption	to	actual utility
compliance	to	excellence

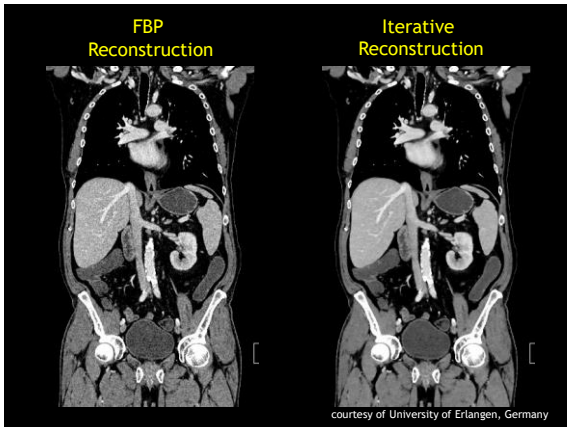
Features of modern CT

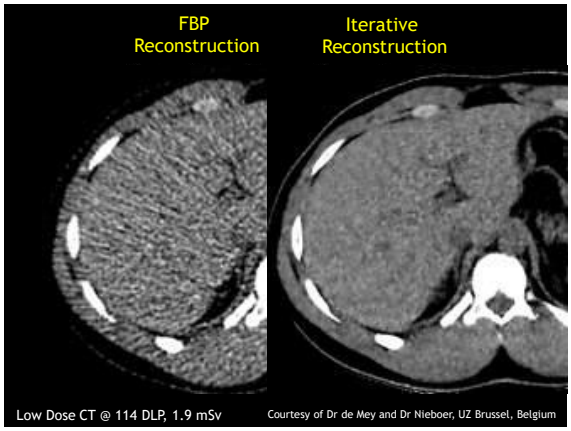
- Variable kernels
- Iterative reconstruction
- Tube current modulation
- Special applications
  - Spectral methods and applications
  - Cone-beam CT
  - Cardiac CT
  - Perfusion imaging

Iterative recons

ASiR (GE)
Veo (GE)
IRIS (Siemens)
SAFIRE (Siemens)
AIDR (Toshiba)
iDose (Philips)

- Significant potential for dose reduction
- Potential for improved image quality
- Increased vendor-dependence
- Unconventional image appearance
- Limited utility of prior quality metrics
- Need for nuanced implementation for effective improvement in patient care

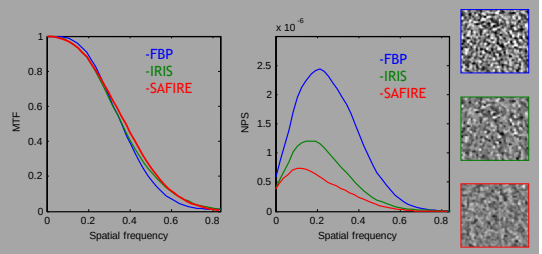




Resolution and noise, eg 1

Comparable resolution

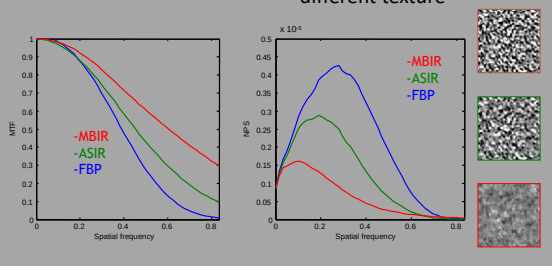
Lower noise but different texture



Resolution and noise, eg 2

Higher resolution

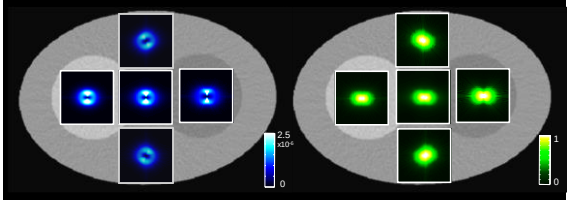
Lower noise but different texture



Non-Stationary Noise and Resolution

The *Local* NPS and MTF

NPS MTF



Noise and resolution model for Penalized Likelihood (PL) model-based reconstruction.* Predictive framework for NPS, MTF, and detectability index (d') enables task-based design and optimization of new systems using iterative reconstruction.

*Fessler et al. IEEE-TIP (1996)
G. Gang et al. Med Phys 41 (2014)

Outline

- Foundation
  - What has led to TG233
- **Methods**
  - **Select procedures**
- Prospects
  - Possibilities
  - Future extensions

Methods

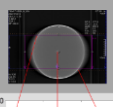
1. Visual inspection
2. Basic performance
   - Summarizing existing methods in tabular form
3. Operational performance
   - Tube current modulation
   - Spatial resolution
   - Noise
   - Quasi-linear task-based performance
   - Spatial domain task-based performance

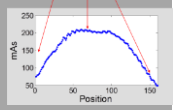
Appendix

Basic performance

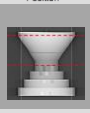
Geometrical performance	Laser alignment accuracy
	Table indexing accuracy
	Slice position accuracy
	Slice thickness accuracy
	Slice thickness accuracy
	Gantry tilt accuracy
Radiation performance	Half-value layer
	Exposure reproducibility
	Timer reproducibility
	Exposure linearity
	kVp accuracy
	Timer accuracy
	Radiation beam profile
	Adaptive collimation
	Accuracy of displayed $CTDI_{vol}$
	Dose of digital survey radiography
Imaging performance	CT number accuracy
	CT number uniformity
	Geometric distortion
	Artifact assessment
	Line-pair resolution
	Noise magnitude
	Low-contrast detectability

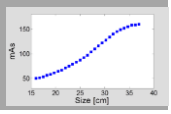
TCM



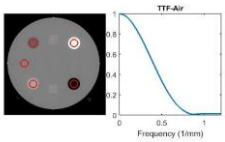


- Use:
  - Variable-sized, continuous and step-wise phantoms imaged with TCM
- Measure:
  - mA and noise per size
- Evaluate:
  - mA and noise size dependencies
  - Phase concordance



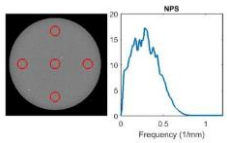


Resolution



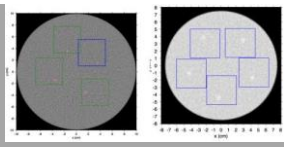
- Use:
  - Phantom with 2 cm circular inserts of relevant contrast imaged using representative protocols
- Measure:
  - Task transfer function (TTF)
- Evaluate:
  - TTF at defined noise and contrast
  - Frequencies at 50% and 10% TTF (f_{50} and f_{10})

Noise



- Use:
  - Variable-sized, uniform phantom imaged using representative protocols
- Measure:
  - Noise magnitude and NPS
- Evaluate:
  - SD and NPS at defined noise levels
  - Peak, average frequencies of the NPS (f_P and f_A)

Spatial-domain task-based detectability



- Use:
  - Uniform phantoms with rod and sphere targets imaged using representative protocols
- Measure:
  - Human or observer model target detection
- Evaluate:
  - Localization success rate for targeted tasks
  - Area under the LROC curve for targeted tasks
  - Area under the EFROC curve for targeted tasks

Quasi-linear task-based detectability index

- Use:
  - TTF and NPS evaluation phantom imaged using representative protocols
- Measure:
  - TTF and NPS
- Evaluate:
  - Detectability indices for reference tasks (1, 5, 10 mm, 10 and 100 HU, designer, rect, Gaussian)

$$\left(d'_{NPWE}\right)^2 = \frac{\left[\iint MTF^2(u,v)W_{Task}^2(u,v)E^2(u,v)dudv\right]^2}{\iint MTF^2(u,v)W_{Task}^2(u,v)NPS(u,v)E^4(u,v) + MTF^2(u,v)W_{Task}^2(u,v)N_i\,dudv}$$

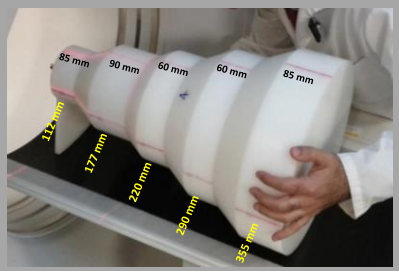
Suggested reference protocols

CTDI	kVp	mA	Mode/Pitch	Reconstruction
0.75	120	Fixed mA to achieve target CTDI ± 10%	Helical, ~1	FBP, IR at mid, upper-mid, and max settings "standard" kernel ~0.6 and 5 mm slice thickness
1.5				
3.0				
6.0				
12.0				
24.0				
3.0	70 (or 80)		Helical, ~1, unless a lower pitch is needed to achieve the CTDI	
3.0	100			
3.0	150 (or 140)			
1.5	120	AEC setting to achieve target CTDI ± 10%	Helical, ~1	
3.0				
6.0				

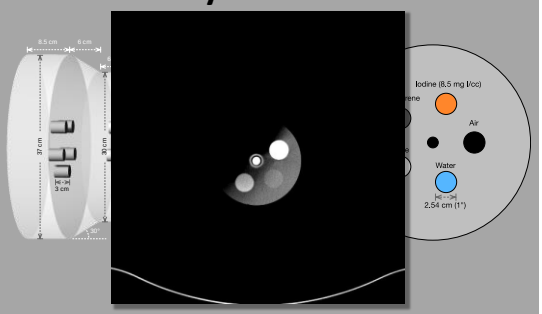
Quasi-linear task-based measurements

Mercury Phantom 3.0

- Size matching population cohorts
- Designed for size, AEC, and d' evaluations



Mercury Phantom 3.0



Design: Size

Pediatric representation percentages

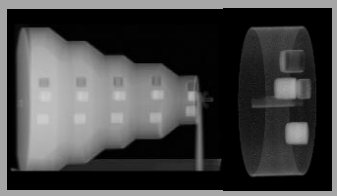
MP 3.0 section	Water size equivalent	Abdomen		Chest		Head	
		Age	Percentile	Age	Percentile	Age	Percentile
120 mm	112 mm	0	12	0	50	0	5
		7	5	5.5	5		
185 mm	177 mm	6	50	10	50	3	95
		15	5	16	5	12	50
230 mm	220 mm	3	95	8	95	-	-
		12	50	16	50		
		21	5				
300 mm	290 mm	12	95	19	95	-	-
		21	50				
370 mm	355 mm	20	95	-	-	-	-

Design: Size

Adult representation percentages

MP 3.0 section	Water size equivalent	Abdomen		Chest		Head	
		M	F	M	F	M	F
120 mm	112 mm	-	-	-	-	-	-
185 mm	177 mm	-	-	-	-	25	75
230 mm	220 mm	0.4	9	0.06	1.4	-	-
300 mm	290 mm	27.1	61	14	48	-	-
370 mm	355 mm	80	90.3	60	87	-	-

Design: Resolution, HU, noise

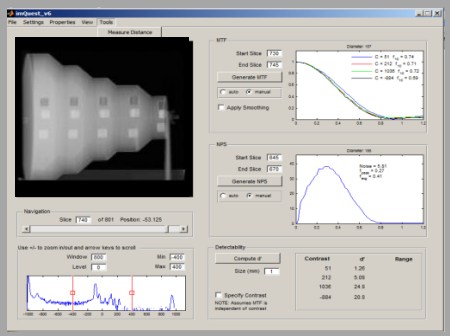


- **Representation of abnormality-relevant HUs**
- **Sizes large enough for resolution sampling**
- **Maximum margin for individual assessment**
- **Iso-radius resolution properties**
- **Matching uniform section for noise assessment**

30

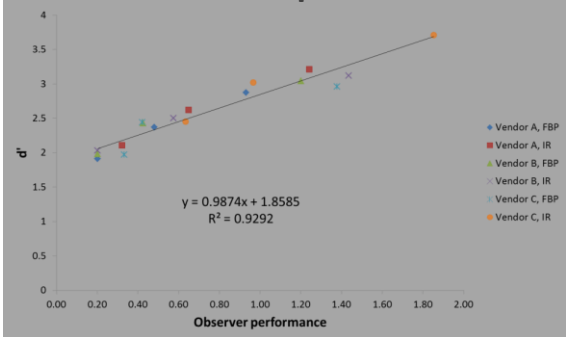
Image Quality Evaluation Software

HU, Contrast, Noise, CNR, MTF, NPS, d' per patient size, fixed mA and TCM



Expected release through AAPM TG 233

d' vs observer performance



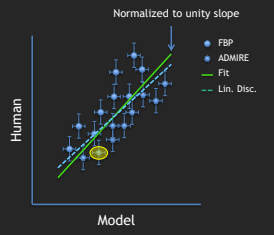
Christianson et al, Radiology, 2015

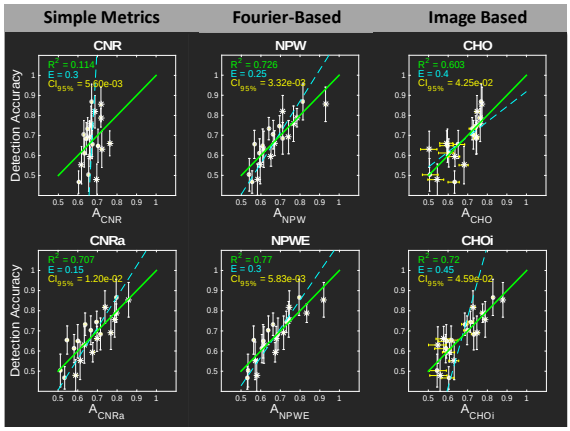
Comparing observer models

Model:	CNR	CNRa	NPW	NPWE	CHO	CHOi
Task Properties						
Lesion Contrast	x	x	x	x	x	x
Lesion Size		x	x	x	x	x
Image Properties						
Noise Magnitude	x	x	x	x	x	x
Noise Texture			x	x	x	x
Resolution			x	x	x	x
Observer Properties						
Visual System				x	x	x
Observer Noise						x
Assumptions						
Quasi LSI System			x	x		
Noise Stationarity			x	x		

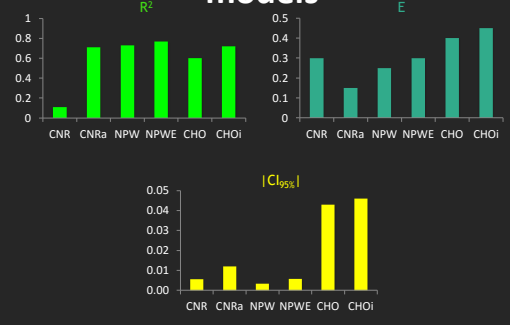
Criteria for goodness of d' calculation methods

- Strong correlation with human results
  - Coefficient of determination (R^2) used as goodness of fit metric.
- Correlation independent of reconstruction algorithm
  - Error, E, of linear discriminator used (bigger = better).
- Confidence interval for d' should be small
  - Average $CI_{95\%}$ used.





Comparing statistics across models



Model Comparison

○ No
◐ Somewhat
● Yes

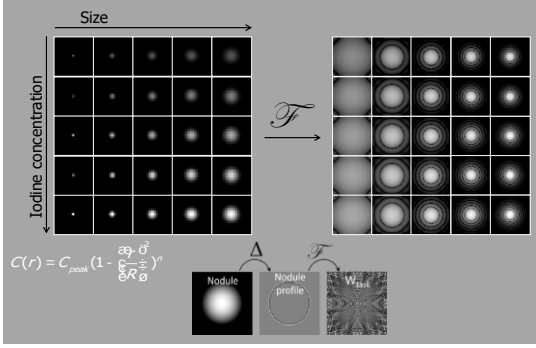
	CNR	CNRa	NPW	NPWE	CHO	CHOi
Correlated with humans?	○	◐	●	●	◐	●
Computed for generic tasks not present in phantom?	○	◐	◐	◐	○	○
Handling non-linearity or non-stationary noise?	○	○	◐	◐	●	●
Correctly characterizing different recons?	○	○	●	●	●	●
Acceptable uncertainty for a reasonable # of images?	●	●	●	●	◐	◐

Outline

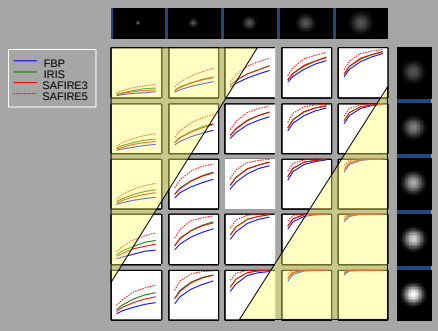
- Foundation
  - What has led to TG233
- Methods
  - Select procedures
- **Prospects**
  - **Possibilities**
  - Future extensions

Task-based performance

Task characteristics for detection and estimation



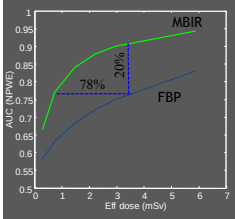
AUC vs. dose



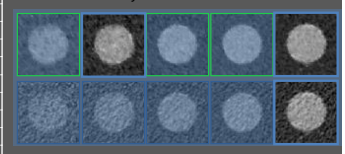
Chen, SPIE, 2013

Task-based dose reduction

Large feature detection task



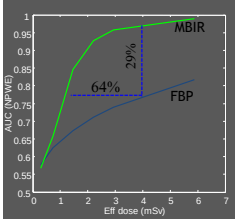
ACR Acrylic insert



0.2 mSv → 3.7 mSv

Task-based dose reduction

Small feature detection task



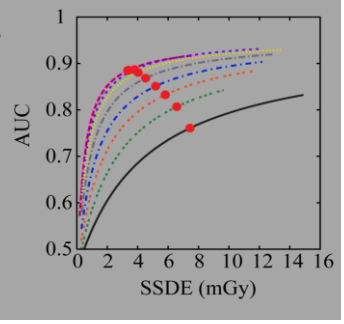
ACR 0.5 lp/mm bar pattern

0.2 mSv → 3.7 mSv

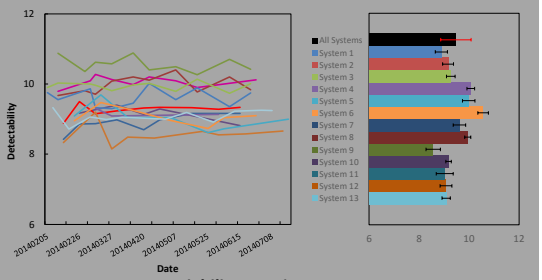
Dose-quality optimization

Quality-dose gradient to achieve highest quality at lowest dose

- Iso-gradient operating points

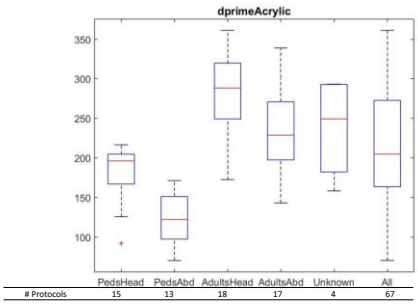


Detectability index across systems



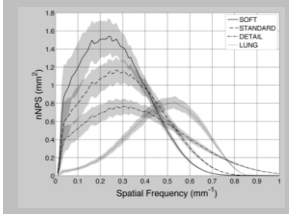
Intra system variability: 1-4%
Inter system variability: 6%

Detectability index across protocols – pilot national trial

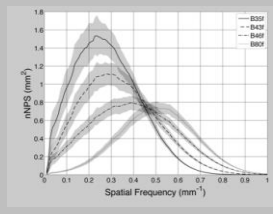


Protocol optimization:
Noise texture matching

GE

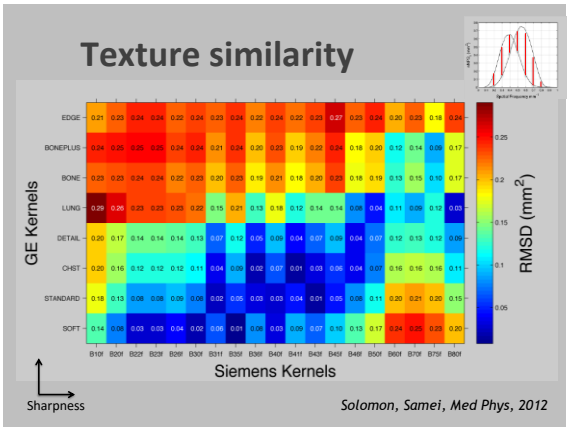


Siemens



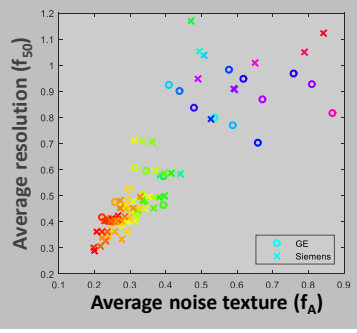
Solomon, Samei, Med Phys, 2012

Texture similarity

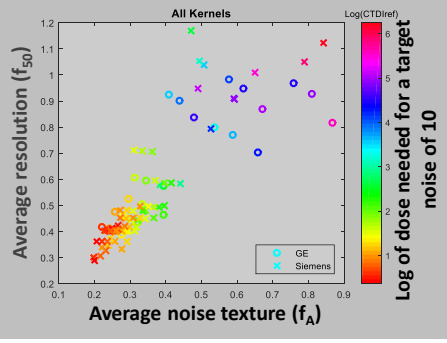


Solomon, Samei, Med Phys, 2012

Protocol optimization:
Noise texture and resolution matching



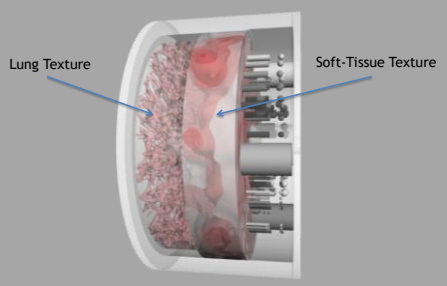
Protocol optimization:
Noise texture and resolution matching

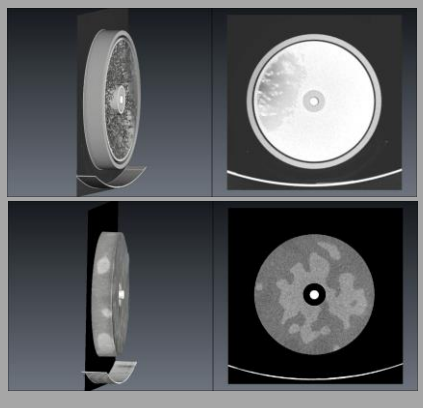


Outline

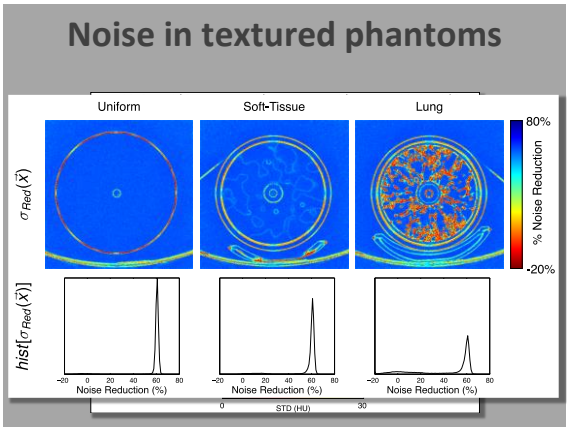
- Foundation
  - What has led to TG233
- Methods
  - Select procedures
- **Prospects**
  - Possibilities
  - **Future extensions**

CT performance in anatomically-informed textured phantoms



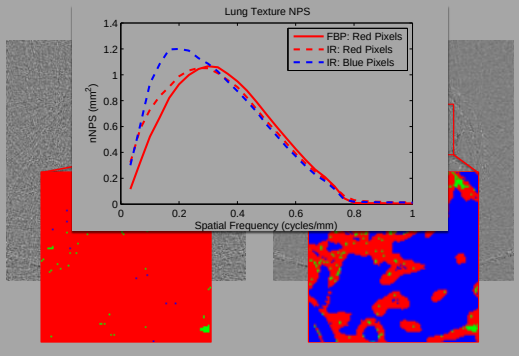


Noise in textured phantoms

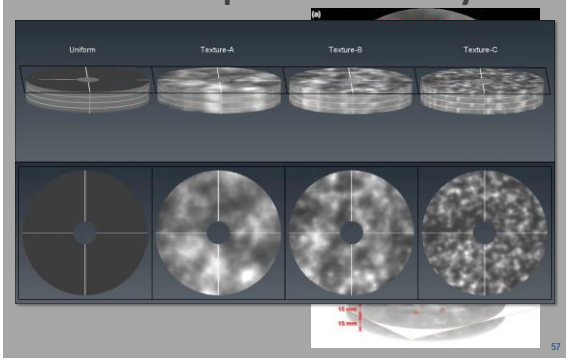


What about noise texture?

$$I(\vec{x}) = \mu(\vec{x}) + N(\vec{x})$$



Texture phantom library



57

Conclusions

- New technologies and new paradigm necessitate an upgrade to performance metrology towards higher degrees of clinical relevance:
  - "Taskful" surrogates of clinical performance
  - Application for use optimization
  - TG233 – a first step towards uniformity and relevance of characterization
- TG 233 timeline
  - Aug 2016: v. 12 released for committee review
  - Release anticipated in late 2016